Title
The challenge of determining appropriate care in the era of patient-centered care and rising health care costs.

Permalink

Journal

ISSN

Authors
Coulter, Ian
Herman, Patricia
Ryan, Gery
et al.

Publication Date

DOI

The challenge of determining appropriate care in the era of patient-centered care and rising health care costs

Ian Coulter[1,2] ⓘD, Patricia Herman[3], Gery Ryan[3], Lara Hilton[4],
Ron D. Hays[5] and Members of CERC Team*

Abstract
Appropriateness of care is typically determined in the United States by evidence on efficacy and safety, combined with the judgments of experts in research and clinical practice, but without consideration of the cost of care or patient preferences. The shift in focus towards patient-centered care calls for consideration of outcomes that are important to patients, accommodation of patient preferences, and incorporation of the costs of care in patient-provider shared clinical decisions. The RAND/UCLA Appropriateness method was designed to determine rates of appropriate or inappropriate care, but the method did not include patient preferences or costs. This essay examines how methods of studying appropriateness can be made more patient-centered by describing a modification of the RAND/UCLA method by including patient outcomes, preferences, and costs.

Introduction

The ultimate goal of health care is to ensure that patients receive appropriate care; that is, the right care for the right patient for the right problem at the right time and from the right provider. Inappropriate care is costly; studies suggest that up to 30% of health care costs are wasted on inappropriate or useless care. The challenge is to define and increase delivery of appropriate care.

In the United States, the 2010 Patient Protection and Affordable Care Act and the establishment of the Patient-Centered Outcomes Research Institute (PCORI) have posed at least two new challenges to the traditional approach to measuring appropriateness. The first is how to include patient preferences into what has largely been clinician-based decision-making. A second challenge is the shift in focus away from explanatory randomized controlled trials (RCTs) to comparative effectiveness research, which seeks to identify what is important to patients, providers, and other stakeholders (effectiveness in everyday practice).[1] The patient wants to know 'will it work for me' and the

provider wants to know 'will it work for my patients', but because of the strict protocols and inclusion and exclusion criteria of RCTs, it is difficult to generalize whether an intervention applies to an individual's circumstances. Furthermore, an increasing share of the costs of care is shifting to patients, which may affect its appropriateness from the patient's perspective. In this essay, we examine how methods of studying appropriateness can be made more patient-centered by

[1]Senior Health Policy Researcher, RAND Corporation, Santa Monica, California, USA
[2]Professor, School of Dentistry, University of California, USA
[3]Senior Behavioral Scientist, RAND Corporation, USA
[4]Senior Research Analyst, RAND Corporation, USA
[5]Professor UCLA Department of Medicine, USA

*The details of the members of the CERC team are given in the Acknowledgements section.

Corresponding author:
Ian Coulter, RAND Corporation Health, 1776 Main Street, Santa Monica, California 90401, USA.
Email: coulter@rand.org

describing a modification of the RAND/UCLA appropriateness method that includes patient outcomes, preferences, and costs. We begin by briefly examining the drivers for an increased interest in assessing appropriate care in the United States and describe the principles of the RAND/UCLA appropriateness method (RAM). We then outline a modification of the RAM, using the example of complementary and alternative medicine. We close with a set of conclusions based on our research.

Drivers for examining appropriate care

In the last decades, much attention has been given to evidence-based practice; that is, care that has been proven efficacious and safe.[2] The implicit assumption was that evidence-based practice ensures the delivery of appropriate care. Yet, perhaps as little as 15%–20% of all medical practice is evidence-based.[3] Hicks[3] notes 'it is generally accepted that between 20% and 60% of patients either receive inappropriate care or are not offered appropriate care.' Moreover, it is not uncommon for evidence to be equivocal or contradictory.[4]

In the United States, an important impetus for examining the appropriateness of care was the observed variation in the amount of medical care delivered that could not be attributed to variations in those receiving it.[5–8] Surgical procedure rates varied widely across geographic areas. Research has established that the incidence and prevalence of disease, socioeconomic factors, and underlying differences in the health care delivery system all contributed to these variations; however, these factors did not adequately explain observed variations.[9] Appropriateness of care was found to be a contributor to geographical variation in the amount of health care delivered.[10]

A second factor driving the interest in appropriateness was the failure of the National Institutes of Health Consensus Development Program, established in 1977, which brought together leading scientists to review the research evidence and to make recommendations for clinical practice, to reach consensus on many indications for treatments. Further research showed that most consensus recommendations had had little impact on the behavior of the practitioners at which they were targeted, thus leaving much practice short of what was recommended.[11]

In response, the RAND Corporation and University of California, Los Angeles (UCLA) developed the RAND/UCLA appropriateness method (RAM). This was designed to have more credibility with clinicians by leveraging their clinical experience to expand the research evidence, address more indications that were relevant to the actual patients they were treating, and not force consensus. RAM is a formal and replicable method that integrates scientific evidence with clinical experience and acumen.[12]

The RAND/UCLA appropriateness method (RAM)

RAM combines clinical expert and researcher panels to consider the available evidence and to judge, for a particular treatment, whether for an average group of patients with a given clinical indication the procedure under consideration is appropriate or not. RAM is widely used for defining and identifying appropriate care in the United States and internationally,[13–15] with other work examining reliability and validity of the method.[16,17]

RAM includes seven steps (Figure 1). It begins with a systematic review of the literature, which includes studies of efficacy and effectiveness.[12,18] Second, based on the review, the researchers and subject matter experts create a set of indications (combinations of patient symptoms, past medical history, and test results) for performing a given clinical procedure or using a drug or supplement. The goal for these indications is to be comprehensive, i.e. they cover all likely
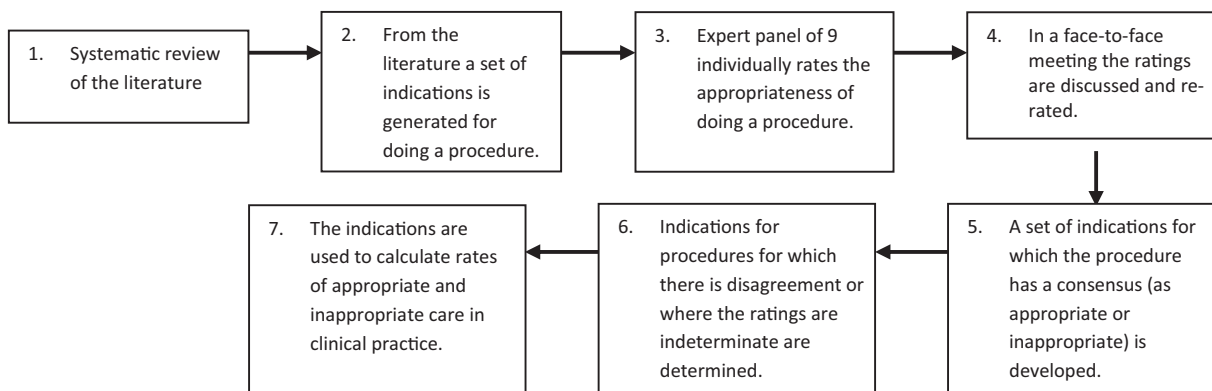
1. Systematic review of the literature →
2. From the literature a set of indications is generated for doing a procedure. →
3. Expert panel of 9 individually rates the appropriateness of doing a procedure. →
4. In a face-to-face meeting the ratings are discussed and re-rated. ↓
5. A set of indications for which the procedure has a consensus (as appropriate or inappropriate) is developed. ←
6. Indications for procedures for which there is disagreement or where the ratings are indeterminate are determined. ←
7. The indications are used to calculate rates of appropriate and inappropriate care in clinical practice.

Figure 1. The process of the appropriateness method.

patient presentations. However, the list of indications must also be short enough that each can be separately and realistically evaluated for appropriateness by the panelists.

Third, a nine-member panel consisting of content experts and clinicians individually rate the appropriateness of the procedure for each indication. Ratings range from 1 to 9, with 1 representing that the procedure is *extremely inappropriate* for that patient type, and 9 that the procedure is *extremely appropriate*, with appropriateness determined by assessing whether 'the expected health benefit to the patient (relief of symptoms, improved functional capacity, reduction of anxiety, etc.) exceeds expected health risks (pain, discomfort etc.) by a sufficiently wide margin that the procedure is worth doing,'[16] i.e. the efficacy/effectiveness and safety of the procedure for each indication. Panelists are asked to evaluate the risks and benefits based on commonly accepted best clinical practice at the time the panel is conducted. This is followed by a face-to-face meeting in which panelists discuss the ratings and are given the opportunity to rerate the indications (step 4). In a fifth step, the expert panel produces a set of indications, based on consensus, for doing or not doing a procedure. It also identifies indications for which the evidence is indeterminate, and indications for which there was no agreement (step 6). The indications for which there is consensus can be used to guide future practice but also as a standard against which to assess current practice. This process makes it possible to calculate the rates of appropriate and inappropriate care (step 7), using randomly selected patient files to assess the care that was provided against the indications.

Going beyond the appropriateness method

RAM makes it possible to take the best of what is known from research, combine it with the expertise of experienced clinicians, and apply it to a wide range of patients and presentations seen in real-world clinical practice. Clinicians are, after all, the final translators of evidence into practice, and RAM formalizes the process. A shortcoming of the approach is that its definition of appropriateness relies heavily on safety, efficacy, and effectiveness. Yet, appropriateness is complex and it requires that care should be 'effective (based on valid evidence); efficient (cost-effective); and consistent with the ethical principles and preferences of the relevant individual, community or society'.[19] Thus, there is room to enhance the methods for defining and identifying appropriate care. We here discuss an advancement to the RAM that explicitly considers patient outcomes, preferences and cost, using the

example of the RAND Center of Excellence for the Study of Appropriateness of Care in Complementary and Alternative Medicine (hereafter, the RAND CERC).

The study of the appropriateness of care in complementary and alternative medicine (CAM)

The RAND CERC focuses on ways to include patient preferences and resource use in determining the appropriateness of chiropractic manipulation and mobilization (M/M) for chronic low back pain (CLBP) and chronic cervical pain (CCP). An estimated 76.5 million adult Americans reported experiencing pain at some point in their life that lasted more than 24 h; of those reporting pain, 42% said it lasted more than a year.[20] The 2010 Global Burden of Disease study found low back pain to be the most common cause of years lived with disability, with neck pain the fourth most common cause.[21] The estimated health care expenditures for adults with spine problems in the United States in 2005 were $85.9 billion, representing 9% of the total national expenditure, with low back pain ranking among the 10 most expensive medical conditions.[22]

The RAND CERC comprises four large-scale research projects focusing respectively on clinicians, outcomes, patient preferences and resource utilization, and two core service modules (a Systematic Review Core and a Data Collection and Management Core). We describe the projects in turn.

Clinician-based appropriateness

This project sought to determine appropriate care for chiropractic M/M for CLBP and CCP. It used two panels of expert clinicians and researchers (CAM and non-CAM) to interpret and synthesize all available data on safety, efficacy, and effectiveness.[12] This approach principally built on earlier work on M/M for acute low back pain,[23] while taking account of the substantial expansion of the evidence base that had evolved since, with review work performed by the Systematic Review Core.[24] Panelists reviewed the literature and rated a comprehensive set of 450 different indications for CLBP and 186 indications for CCP.[24,25] The main objectives were to develop these ratings and to apply them to a national sample of patients' charts (collected through the Data Collection and Management Core; Box 1) to determine the prevalence of appropriate and inappropriate care in practice.[24] The two expert panels were reconvened following completion of the three other projects described below, which provided patient-centered information on

> **Box 1. Data collection and management core.**
>
> All four projects drew on a single data core, using a national cluster sample to recruit chiropractic practices across six states: California, Florida, Minnesota, New York, Oregon, and Texas. The states were chosen to ensure (i) geographical representation of the major areas of the United States and (ii) variation in chiropractic practice across the USA. In each state a single metropolitan center was selected and approximately 20 chiropractic clinics were recruited. In each practice patients were recruited over a four-week period using an onsite iPad; eligible patients were asked to complete online surveys at baseline and three months, and shorter surveys every two weeks in -between. We successfully recruited 2024 patients from 125 clinics across the six states; of patients who completed the baseline, 91% also completed the 3-month survey and respondents completed on average four of the five bi-weekly surveys.
>
> We further scanned the files of a random retrospective sample of patients at each clinic to determine the amount of chronic pain being treated in a given clinic. The data from these files were then abstracted, and we are in the process of comparing these data to our appropriateness ratings to determine the overall rates of appropriate and inappropriate care.

outcomes, preferences, and resource utilization. Panel members were asked to rerate the appropriateness of M/M for each clinical indication to determine if patient-centered information led the panel members to change their ratings.

The panels' ratings provide feedback to the research community regarding patient presentations to a health care provider where substantial uncertainty remains as to the appropriateness of M/M, and feedback as to whether guidelines are needed to increase appropriate and reduce inappropriate care.

Outcomes-based appropriateness

The second project examined the relationship of appropriateness of care with patient experiences of care and health-related quality of life. It aimed to adapt existing state-of-the-science measures for patients using M/M for CLBP and CCP and evaluate the relationships of appropriate care with these patient-reported measures, drawing on systematic review evidence on outcome measures performed by the Systematic Review Core and using our national sample of patients' charts (Box 1). We also conducted focus groups and cognitive interviews to help evaluate candidate survey instruments and develop supplemental items meaningful to patients and included domains relevant to them.

Patient preference-based appropriateness

Given the prevalence of patient self-referral and the health system-wide focus on patient-centered care, this project examined how patient preferences affect what is considered appropriate care. Objectives for this study included understanding how CLBP and CCP patients decide to use M/M and determining what they believe is appropriate care, again drawing on systematic review evidence on measures of patient preferences performed by the Systematic Review Core

and data from our national sample (Box 1) to determine patient preferences for M/M care.

Resource utilization-based appropriateness

Finally, given the high proportion of out-of-pocket costs in CAM and the rise of health care costs in general, this project examined how cost affects the type of care that is most appropriate and the appropriate course of care once a type of care has been identified. It aimed to determine the relative cost-effectiveness of care options for CLBP and CCP using two economic simulation models, with related evidence reviews supporting model building performed by the Systematic Review Core and drawing on longitudinal data from our national sample (Box 1) to understand the characteristics of an appropriate course of care once M/M is selected.

Conclusions

In this essay, we have described a method for incorporating a patient-centered perspective with its focus on shared decision-making in health care in traditional clinically based appropriateness of care studies in CAM. The methods we have developed are applicable for other types of CAM and other conditions, and for health care in general. But, as this essay suggests, it poses considerable implementation challenges. On the one hand, the care patients receive should be clinically appropriate and clinically necessary. This will involve clinical judgments, which, for the most part, patients cannot make unassisted and judgments by clinicians based on evidence, experience, and clinical acumen will remain necessary. But these judgments should be and hopefully will be increasingly strengthened by patient input. The challenge is how to make sure the judgments do not conflict. Providers are expected to

understand patients' needs and wants and encouraged to advise patients accordingly.

Patients' preferences and financial burden should also be considered as criteria in the clinical decision making once safety and efficacy/effectiveness have been established. In this era of rising health care costs, it is increasingly important to evaluate the appropriateness of therapies provided to patients. This is even more the case with CAM therapies, as these remain less well-researched, even as their utilization remains high.

Crowdsourcing as a supplement to appropriateness methods

One of the major limitations facing including patients' preferences and costs is that this adds further costs to appropriateness studies, which are already considerable. The RAND CERC was funded by U.S. National Institutes of Health (NIH) at a cost of $8.5 million. Although the study was successful in recruiting clinics and patients (Box 1), collecting practice-based data is both costly and time consuming.

One new method, which promises to greatly reduce the cost, is crowdsourcing and it was used in this project as an alternative approach to collect patient data. *Crowdsourcing* is defined as the paid recruitment of an online, independent, global workforce for the objective of working on a specifically defined task or set of tasks.[26] It first emerged in 2006 as an online contract labour market where needed services, ideas, or content were obtained for pay from a large group of people.[27] More recently, platforms such as Amazon Mechanical Turk have been tested for the use for medical, psychological, and behavioral sciences research. This was shown to provide data faster and more cheaply as well as affording access to a broader cross-section of the population than is typical of standard research experiments.[28–30]

While providing potential for low-cost recruitment and data collection for clinical practice, crowdsourcing has not been validated as a data collection tool for chronic pain research. We used the RAND CERC data by comparing patient-reported pain, function, quality of life, and demographics to data collected from MTurk participants who seek chiropractic treatment for their CLBP. Our data show it is a cost-efficient way to collect data and generates a patient sample that is not significantly different from that we collected in the practices.

ORCID iD

Ian Coulter http://orcid.org/0000-0002-2352-1939

References

1. Coulter ID. Comparative effectiveness research: does the emperor have clothes? *Alt Ther Health Med* 2011; 17: 8–15.
2. Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: clinical practice guidelines. *BMJ* 2016; 353: i2089.
3. Hicks NR. Some observations on attempts to measure appropriateness of care. *BMJ* 1994; 309: 730.
4. Coulter ID, Lewith G, Khorsan R, et al. Research methodology: choices, logistics, and challenges. *Evid Based Complement Alt Med* 2014; 2014: 780520.
5. Chassin MR, Brook RH, Park RE, et al. *Variations in the use of medical and surgical services by the medicare population*. RAND, Santa Monica USA, 1986.
6. Wennberg JE and Gittelsohn AM. *Small area variations in health care delivery*. Washington, DC: American Association for the Advancement of Science, 1973.
7. Lewis CE. Variations in the incidence of surgery. *N Engl J Med* 1969; 281: 880–884.
8. Roos NP. Hysterectomy: variations in rates across small areas and across physicians' practices. *Am J Public Health* 1984; 74: 327–335.
9. Wennberg JE. The paradox of appropriate care. *JAMA* 1987; 258: 2568–2569.

10. Chassin MR, Kosecoff J, Park RE, et al. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *JAMA* 1987; 258: 2533–2537.

11. Lomas J. Words without action? The production, dissemination, and impact of consensus recommendations. *Annu Rev Public Health* 1991; 12: 41–65.

12. Coulter I, Elfenbaum P, Jain S, et al. SEaRCH™ expert panel process: streamlining the link between evidence and practice. *BMC Res Notes* 2016; 9: 16.

13. Andreasen B. Consensus conferences in different countries. *Int J Technol Assess Health Care* 1988; 4: 305–308.

14. McClellan M and Brook RH. Appropriateness of care: a comparison of global and outcome methods to set standards. *Med Care* 1992; 30: 565–586.

15. Coulter ID. Evidence-based practice and appropriateness of care studies. *J Evid Based Dent Pract* 2001; 1: 222–226.

16. Shekelle PG, Kahan JP, Bernstein SJ, et al. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med* 1998; 338: 1888–1895.

17. Coulter I, Adams A and Shekelle P. Impact of varying panel membership on ratings of appropriateness in consensus panels: a comparison of a multi-and single disciplinary panel. *Health Serv Res* 1995; 30: 577.

18. Coulter ID. Expert panels and evidence: the RAND alternative. *J Evid Based Dent Pract* 2001; 1: 142–148.

19. World Health Organization. Appropriateness in health care services: report on a WHO workshop, Koblenz, Germany 23–25 March 2000. European Health 21 Target 18.

20. Nahin RL, Boineau R, Khalsa PS, et al. Evidence-based evaluation of complementary health approaches for pain management in the United States. *Mayo Clin Proc* 2016; 91: 1292–1306.

21. Hurwitz EL, Randhawa K, Yu H, Cote P, Haldeman S. The Global Spine Care Initiative: a summary of the global burden of low back and neck Pain studies. *Eur Spine J* 2018; (Suppl 6): 796–801.

22. Walsh NE, Brooks P, Hazes JM, et al. Standards of care for acute and chronic musculoskeletal pain: the Bone and Joint Decade (2000–2010). *Arch Phys Med Rehab* 2008; 89: 1830–1845.

23. Shekelle PG, Coulter I, Hurwitz EL, et al. Congruence between decisions to initiate chiropractic spinal manipulation for low back pain and appropriateness criteria in North America. *Ann Intern Med* 1998; 129: 9–17.

24. Coulter I, Whitley M, Hurwitz E, et al. *Determining the appropriateness of spinal manipulation and mobilization for chronic low back pain: indications and ratings by a multidisciplinary expert panel. Report no. RR2475.* Santa Monica, CA: RAND Corporation, 2018.

25. Coulter I, Whitley M, Vernon H, et al. *Determining the appropriateness of spinal manipulation and mobilization for chronic neck pain. Indications and Ratings by a Multidisciplinary Expert Panel.* Santa Monica, CA, USA, 2018.

26. Behrend TS, Sharek DJ, Meade AW, et al. The viability of crowdsourcing for survey research. *Behav Res Meth* 2011; 43: 800–813.

27. Howe J. The rise of crowdsourcing. *Wired Magazine* 2006; 14: 1–4.

28. Hays RD, Liu H and Kapteyn A. Use of Internet panels to conduct surveys. *Behav Res Meth* 2015; 47: 685–690.

29. Weiner M. The potential of crowdsourcing to improve patient-centered care. *Patient* 2014; 7: 1237.

30. Mason W and Suri S. Conducting behavioral research on Amazon's Mechanical Turk. *Behav Res Meth* 2012; 44: 1–23.